BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images, please do not report the images to the Image Problem Mailbox.

31. (Twice Amended) The peptide of claim 1, wherein the peptide contains at least one D-amino acid substitution.

B10 D1

32. (Twice Amended) The peptide of claim 1, wherein the peptide contains at least one variant amino acid substitution selected from the group comprising C_a-methylamino acids, N_a-methylamino acids and α,β-unsaturated amino acids.

33. (Twice Amended) The peptide of claim 1, wherein the peptide is cyclized.

B11

61. (Amended) A purified peptide comprising the amino acid sequence of SEQ ID NO:18.

Pursuant to 37 CFR 1.121(c)(1)(ii), a marked up version of the claims showing the changes made appears as Appendix B of this Amendment.

In the Drawings:

Please enter enclosed formal FIGS 1-16 in replacement for the formal drawings of record.

REMARKS

Upon entry of the amendments made herein, claims 1-18, 31-34, 39, 43-44, 46-47, 49-55, and 57-61 are currently pending in the case. Amendments to the legends for FIGS. 2, 3, 4, 6, 10, 11 and 13, on page 11, line 21, through page 12, line 21, reflect the modified FIG. labels in the replacement formal drawings submitted herewith, as required under 37 CFR § 1.84(u)(1) and in the Office Action. Support for the amendment on page 48, line 7, of the specification can be found in the specification on line 78, line 9.

Claims 45, 48 and 56 have been cancelled without prejudice or disclaimer as they are drawn to a nonelected species. Claims 4 and 6 have been amended to remove recitation of "may be," with amendments supported by claim 1 as originally filed. Claims 12 and 61 have been amended to recite amino acid "sequences" rather than "residues", with amendments supported in the specification at least, *e.g.*, at page 19, lines 11-12. Claims 18 and 31-33 have been amended

to correctly refer to a single peptide of claim 1, with amendments supported by claim 1 as originally filed. No new matter has been added.

Applicants' election of the species encompassed by SEQ ID NO:18 is noted in the Office Action. However, Applicants respectfully dispute the Examiner's assertion (*see* page 2, paragraph 2 of the Office Action) that no allowable generic or linking claims exist pursuant to the nonelected species. At the very least, SEQ ID NO:2 and SEQ ID NO:18 are mouse and human homologs of the elected peptide of the invention, and the generic sequence of claim 4 encompasses the same. Support for this characterization appears in the specification, *e.g.*, at least on page 8, lines 9-13, and in FIG. 16. Applicants believe this interpretation to be correct, and their comments below are presented accordingly.

Objections and rejections have been applied to the specification and pending claims. Each will be addressed in turn, below.

Drawings:

The Examiner has required that FIGS. 2, 3, 4, 6, 10, 11 and 13 be amended to meet the requirements under 37 C.F.R. §1.84(u)(1). Corrected final drawings are submitted herewith. Corresponding amendments have also been made to correct the "DESCRIPTION OF THE FIGURES" section.

Objection to the Specification:

The Examiner states that the specification was objected to because of an informality on page 48, line 7, characterized as needing a sequence identifier. The specification has been amended to contain the appropriate sequence identifier. Applicants respectfully submit that this objection is now moot and request that it be withdrawn.

Rejections under 35 U.S.C. §112, first paragraph

Claims 1-3, 7, 8-18, 31-34 and 39 were rejected by the Examiner under 35 U.S.C. §112, first paragraph, for encompassing purified leptin "homologs, analogs and derivatives," terms which the Office Action characterizes as not sufficiently described in the specification.

Applicants traverse this rejection as applied to the claims as now pending on the grounds that these are common terms known to those skilled in the chemical arts and are clearly given their ordinary meaning within the specification.

Long-standing dictionary definitions are provided herewith for "homologue" (Exhibit 1), "homologous "(Exhibit 2), "analog" (Exhibit 3), and "derivative" (Exhibit 4). Applicants note that these terms, given their ordinary meaning, reasonably convey the full scope of the subject matter of invention to one skilled in the art, and meet all written description requirements. In addition, specific examples of each variant type are disclosed and supported in the specification. Homologs are supported at least, *e.g.*, on page 8, lines 9-13, disclosing SEQ ID NOS:2 and 18 as mouse and human variants. Analogs are described and supported at least, *e.g.*, on page 56, line 6, through page 57, line 9, disclosing cyclization of leptin peptides. Derivatives are described and supported at least, *e.g.*, on page 55, line 14, through page 56, line 5, disclosing D-amino acid substituted leptin peptides. Specifically contemplated peptide designs are described at least on page 54, line 17, through page 58, line 4. Applicants respectfully submit that the specific and even general disclosure relating to the terms "homolog", "homologous", "analog", and "derivative" is amply sufficient to meet the requirements of 35 U.S.C. §112, first paragraph. Applicants therefore respectfully submit that withdrawal of the rejection is proper and in order.

Claims 1-4, 6-18, 31-34, 39, 45, 48, 56 and 61 were rejected under 35 U.S.C. §112, first paragraph, as allegedly lacking enablement. Specifically, the Examiner argues that the specification is enabling for a leptin fragment comprising the amino acid sequence of SEQ ID NO:2 or 18, but leptin peptides encompassing variations in % homology, amino acid substitutions, derivatives, etc., are not enabled. Applicants traverse this rejection as applied to the claims as now pending.

As stated above, the specification on page 54, line 17, through page 58, line 4, clearly discloses cyclized leptin peptides and D-amino acid substituted leptin peptides, among others. One of ordinary skill in the chemical art will clearly recognize that such a modified leptin peptide is defined by the dictionary definition of derivative, as either "a chemical substance related structurally to another substance and theoretically derivable from it" or as "a substance that can be made from another substance." (Exhibit 4) On page 55, lines 24-27, derivatized leptin

peptides are explicitly described where, *e.g.*, "series of peptide analogs will be synthesized, each of which will contain a single D-amino acid which corresponds to its L-isomer in the native sequence."

Determination of percent homology has been routinely done for decades, and is common knowledge for a person skilled in the art. See, *e.g.*, Needleman and Wunsch (1970) *J. Mol Biol.* 48: 443-453 (Exhibit 5). As an example, a skilled artisan will know to compare the mouse and human peptide homologs of SEQ ID NOS:2 and 18, determine that three out of seven residues differ, and calculate that the remaining four out of seven residues compute to a percent homology of 57%. Simple mathematical modifications from this formula, widely practiced by those skilled in the art, are likewise used to determine the number of residues that can differ, *e.g.* in a seven residue peptide (SEQ ID NOS:2 and 18) or a fifteen residue peptide (SEQ ID NOS:3-10) that will correspond to a polypeptide with at least 70% homology, or greater than 85% or 95% homology, to such peptide.

As for the case law cited by the Examiner, the disclosures at issue within those cases are not appropriate to the present invention. For instance, the passage of *in re Fisher* is cited for the proposition that an inventor not be allowed to obtain broad claim scope in the absence of "reasonable correlation" to the scope of enablement in the specification. In the present case, Applicants have provided specific peptides with leptin activity, and have explicitly described methods of preparation, physical or chemical properties, and other characteristics sufficient to define a leptin peptide of the invention. See, *e.g.*, the specification at least at page 54, line 17, through page 58, line 4. The reference to *Amgen* is similarly inapplicable, where the disclosure in *Amgen* lacked even a "mental picture of the structure of the chemical." See, Amgen Inc. v. Chugai Pharmaceuticals Co. Ltd., 18 U.S.P.Q.2d, 1016. Applicants' claimed invention requires, *inter alia*, leptin peptides having body mass modulating ability, while retaining strong homology to specifically recited peptide sequences. In addition, and as stated in the paragraph above, determination of homologs, analogs and derivatives of peptides of, or percent homology to, known sequence is well-known and commonly practiced in the art, and therefore is within the skill of those of ordinary skill in the art. For instance, at the time this application was filed, leptin peptides from many different species were already cloned and sequenced, and determined to be homologs across many species. See, *e.g.*, Appendix C, wherein leptin clones publicly

disclosed prior to August 1999 are aligned and identified by GenBank accession number and by genus species designation. It therefore follows that reference to *Clark* is also inappropriate where the Examiner has specifically acknowledged disclosure by the Applicant of at least two species (murine and human) in the specification, and many additional species were provided by those skilled in the art at the time of the invention. See, *e.g.*, Appendix C.

In summary, Applicants believe reference to at least at page 54, line 17, through page 58, line 4, in the specification is sufficient enablement under 35 U.S.C. §112, first paragraph.

Applicants submit that these 35 USC § 112, first paragraph, rejections are moot or overcome as described above, and request that these rejections be withdrawn.

Rejections under 35 U.S.C. §112, second paragraph

Claims 4, 6, 8-12, 18, 31-33, 45, 48, 56 and 61 were rejected under 35 U.S.C. §112, second paragraph as indefinite for reciting various phrases the Examiner found objectionable. Applicants believe these rejections are moot as applied to the claims as now pending. Each phrase rejection will be addressed in turn.

Claims 4 and 6 were rejected for reciting "may be" instead of the more definite verb "is." Claims 4 and 6 have been amended as suggested by the Office Action.

Claims 8-11 and 45 were rejected for reciting "mammalian", "murine", "human", and "synthetic" without distinguishing one from another. Claim 45 has been withdrawn from consideration. Applicants traverse the rejection as applied to claims 8-11 on the grounds that these are common terms known to those skilled in the chemical arts and are clearly given their ordinary meaning within the specification.

Dictionary definitions, which also note the year each term began to be used in the English language, are provided herewith for "mammal" (Exhibit 6) as relating to "mammalian", "murine" (Exhibit 7), "human" (Exhibit 8), and "synthetic" (Exhibit 9). Applicants respectfully submit that these terms, given their ordinary meaning, particularly point out and distinctly claim the subject matter of the invention, as required by 35 U.S.C. §112. As a secondary matter, each of claims 8-11 depend separately from claim 1. In contrast to what is implied in the Office Action on page 8, last paragraph, these cited terms need not be mutually exclusive, so it is not necessary to distinguish one from the other, where a peptide can indeed be both human and synthetic. A

peptide that contains amino acid sequences common to both the murine and human peptides that meets the requirements of claim 1, an independent claim, is still encompassed within the invention without needing to fall within the narrower subsets of dependent claims.

Claims 12 and 61 were rejected for reciting a peptide comprising amino acids (or amino acid residues), stating that the claims are confusing because it is not clear if the peptide has a specific amino acid sequence, or if the peptide just comprises the amino acid in the recited sequence in any given order. Claims 12 and 61 have been amended to more clearly distinguish the subject matter of the claims. This rejection is now moot as applied to the claims as amended.

Claims 18, 31, 32 and 33 were rejected for reciting "any one of the peptides of claim 1" or similar language, where claim 1 is limited to a single peptide. Claims 18, 31, 32 and 33 have been amended to clarify the claim. This rejection is now moot as applied to the claims as amended.

Claims 45, 48 and 56 are rejected for depending from a non-elected claim, where the claim was withdrawn for not reading on the elected species. Claims 45, 48 and 56 have been withdrawn, so the rejection is now moot.

These 35 USC § 112, second paragraph, rejections are believed moot or inappropriate as applied to the claims as now pending. Applicants request that these rejections be withdrawn.

Rejections under 35 U.S.C. §102(a) and 35 U.S.C. §102(b) over Grasso et al.

Claims 1-4, 6-18, 39, 45 and 61 were rejected as anticipated by Grasso *et al.* (*Endocrinol.* 138: 1413-1418, 1997) ("Grasso"). Applicants traverse this rejection.

The leptin fragments described in Grasso are identified in the specification as SEQ ID NOS: 11-16. The claims are drawn to polypeptides of SEQ ID NOS: 2-10 and 18, where SEQ ID NO:18 is the elected peptide currently under consideration. It is settled law that a reference that does not describe each and every element of a claimed invention cannot be a bar to patentability under 35 U.S.C. §102. See, Hybritech Inc. v. Monoclonal Antibodies, Inc. 802 F.2d 1367, 231 USPQ 81 (Fed. Cir. 1986). At the very least, the 15 amino acid peptides in Grasso are more than twice the length of the 7 amino acid polypeptide of SEQ ID NO:18, and Grasso does not provide the leptin polypeptide of SEQ ID NO:18. Therefore, Grasso is not prior art to the present invention. Applicants request that these rejections be withdrawn.

Rejections under 35 U.S.C. §102(b) over Samson et al.

Claims 1-3, 9-14, 16, 18, 33 and 39 were rejected as anticipated by Samson *et al.* (*Endocrinol.* 137(11): 5182-5185, 1996) ("Samson"). Applicants traverse this rejection.

The three leptin fragments described in Samson are from 35 amino acids to 52 amino acids in length, in comparison to the 7 amino acid polypeptide of SEQ ID NO:18. Since Samson et al. does not describe each and every element of the claimed invention, Applicants request that the rejection be withdrawn.

Rejections under 35 U.S.C. §102(b) over Al-Barazanji et al.

Claims 1-4, 6-18, 39 and 61 were rejected as anticipated by Al-Barazanji *et al.* (PCT Publication WO 97/46585, published 12/11/97) ("Al-Barazanji"). Applicants believe this rejection is moot as applied to the claims as now pending.

Al-Barazanji fails to described any leptin fragments claimed in the invention, and particularly does not disclose the polypeptide of SEQ ID NO:18. Since Al-Barazanji does not describe each and every element of the claimed invention, Applicants request that the rejection be withdrawn.

CONCLUSION

On the basis of the foregoing amendments, Applicants respectfully submit that the pending claims are in condition for allowance. If there are any questions regarding these amendments and remarks, the Examiner is encouraged to contact either of the undersigned at the telephone number provided below. The Commissioner is hereby authorized to charge any underpayments, or credit any overpayment of same, to Deposit Account No. 50-0311 (Reference No. 19705-001).

Respectfully submitted,

Dated: February 19, 2002

Ivor R. Elrifi, Reg. No. 39,529
Kristin E. Konzak, Reg. No. 44,848
Attorneys/Agents for Applicants
MINTZ, LEVIN, COHN, FERRIS,
GLOVSKY and POPEO, P.C.
One Financial Center
Boston, Massachusetts 02111
Tel: (617) 542-6000
Fax: (617) 542-2241

Appendix A: marked up version of the specification showing the changes made

In the Specification:

Replace the paragraphs on page 11, lines 21-26, with the following paragraphs:

[FIG. 2 is a graphic representation] FIGS. 2A-2B are graphic representations of the effects of synthetic leptin peptide OB-3 on body weight gain and food intake in genetically obese female C57BL/6J *ob/ob* mice.

[FIG. 3 is a graphic representation] FIGS. 3A-3B are graphic representations of the effects of synthetic OB-3 on body weight gain and food intake in genetically obese female C57BLKS/J-m *db/db* mice.

[FIG. 4 is a graphic representation] FIGS. 4A - 4N are graphic representations of the effects of 7 daily injections of various synthetic leptin peptides [(**Panels A - N**)] on body weight gain in female C57BL/6J *ob/ob* mice.

Amend the paragraph on page 12, lines 3-4, as follows:

[FIG. 6 is a graphic representation] FIGS. 6A - 6D are graphic representations of the effects of 12 daily injections of various synthetic leptin peptides [(**Panels A - D**)] on body weight gain in female C57BL/6J *ob/ob* mice.

Amend the paragraphs on page 12, lines 12-16, as follows:

[FIG. 10 is a graphic representation] FIGS. 10A - 10B are graphic representations of the effects of 12 daily injections of LEP(116-130) synthetic peptide on body weight gain and food consumption in female *db/db* mice.

[FIG. 11 is a graphic representation] FIGS. 11A - 11B are graphic representations of the effects of 7 daily injections of various synthetic leptin peptide on body weight gain [(**Panel A**)] and food consumption [(**Panel B**)] in genetically obese female C57BLKS/J-m *db/db* mice.

Amend the paragraph on page 12, lines 19-21, as follows:

[FIG. 13 is a graphic representation] FIGS. 13A - 13B are graphic representations of the effects of LEP(116-130) peptide on thermogenesis in genetically obese female C57BLKS/J-m *db/db* mice after 4 days [(**Panel A**)] and 7 days [(**Panel B**)] of peptide treatment..

On page 48, line 7, immediately before the word "was" please insert --(SEQ ID NO:2)--. On line 12, delete "waster" and insert --water--.

In the Drawings:

Delete previously filed FIGS 1-16 and replace them with the enclosed formal drawings FIGS 1-16.

Appendix B: marked up version of the claims showing the changes made

In the Claims:

Please cancel claims 45, 48 and 56 without prejudice or disclaimer.

Please amend claims 4, 6, 12, 18, 31-33 and 61 as follows:

--4. (Twice Amended) A leptin peptide having the amino acid sequence Xaa_n-Ser-Cys-Xaa_1-Leu-Pro-Xaa_2-Xaa_3-Xaa_n, wherein:

(a) either Xaa_n [may be] is zero or a contiguous stretch of at most seven peptide residues derived from SEQ ID NOS: 1 or 17; and

(b) Xaa_1, Xaa_2 and Xaa_3 [may be] is any amino acid substitution. --

--6. (Twice Amended) The leptin peptide of claim 4, wherein:

(a) Xaa_1 [may be] is selected from the group consisting of His or Ser;

(b) Xaa_2 [may be] is selected from the group consisting of Trp or Gln;

(c) Xaa_3 [may be] is selected from the group consisting of Ala or Thr; or

(d) the leptin peptide contains any combination of (a) or (b) or (c). --

--12. (Twice Amended) A peptide comprising an amino acid [residues] sequence of the leptin protein of any one of SEQ ID NOS:1 and 17, selected from the group consisting of:

(i) a sequence comprising amino [acid residues] acids 21-35 (SEQ ID NO:3);

(ii) a sequence comprising amino [acid residues] acids 31-45 (SEQ ID NO:4);

(iii) a sequence comprising amino [acid residues] acids 41-55 (SEQ ID NO:5);

(iv) a sequence comprising amino [acid residues] acids 51-65 (SEQ ID NO:6);

(v) a sequence comprising amino [acid residues] acids 61-75 (SEQ ID NO:7);

(vi) a sequence comprising amino [acid residues] acids 71-85 (SEQ ID NO:8);

(vii) a sequence comprising amino [acid residues] acids 81-95 (SEQ ID NO:9);

(viii) a sequence comprising amino [acid residues] acids 91-105 (SEQ ID NO:10);

(ix) a sequence comprising mouse amino [acid residues] acids 116-122 (SEQ ID NO:2);

(x) a sequence comprising human amino [acid residues] acids 116-122 (SEQ ID NO:18);

and fragments, derivatives, homologs and analogs thereof. --

-- 18. (Twice Amended) A [pharmaceutical] composition comprising [any one of the leptin peptides] the leptin peptide of claim 1, and a pharmaceutically acceptable carrier. --

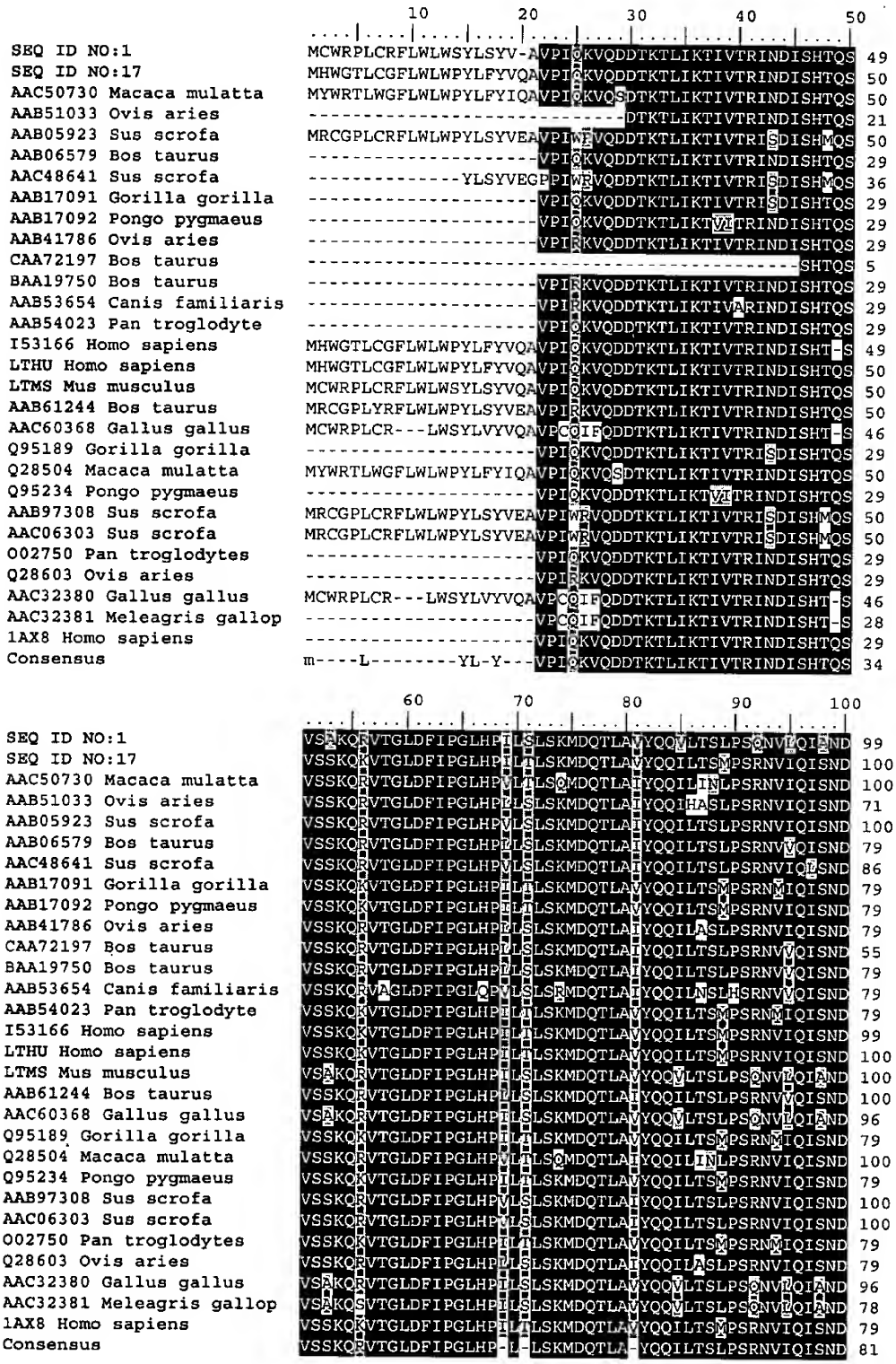
--31. (Twice Amended) [Any one of the peptides] The peptide of claim 1, wherein the peptide contains at least one D-amino acid substitution.

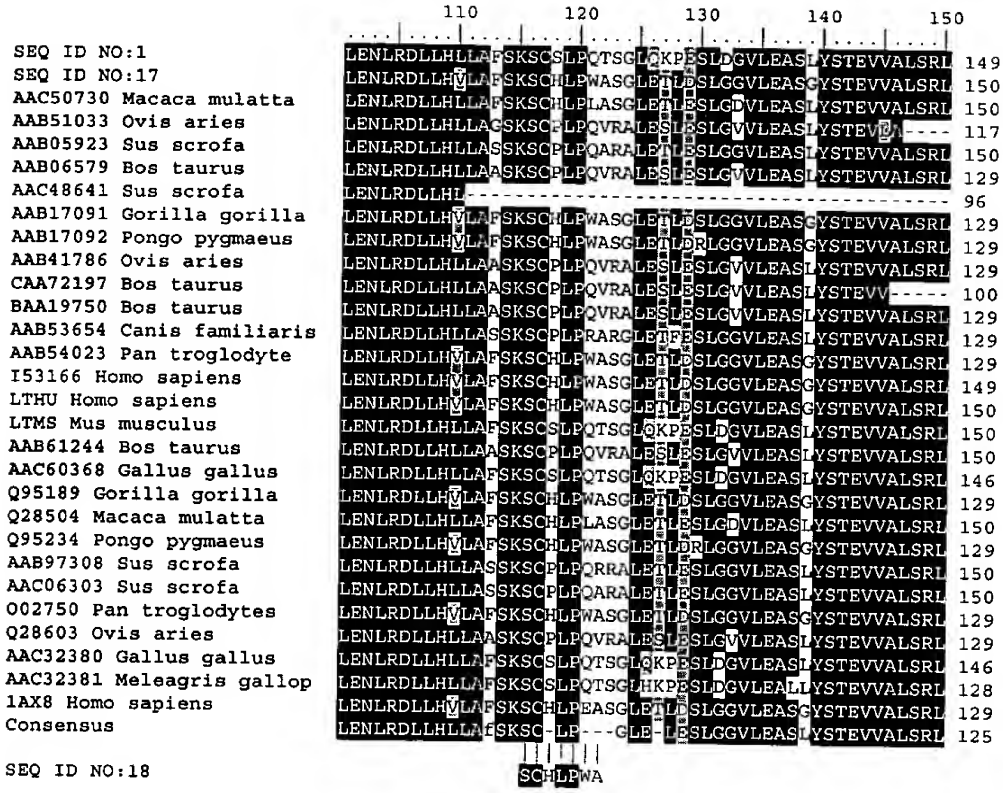
--32. (Twice Amended) [Any one of the peptides] The peptide of claim 1, wherein the peptide contains at least one variant amino acid substitution selected from the group comprising C_a-methylamino acids, N_a-methylamino acids and α,β-unsaturated amino acids.--

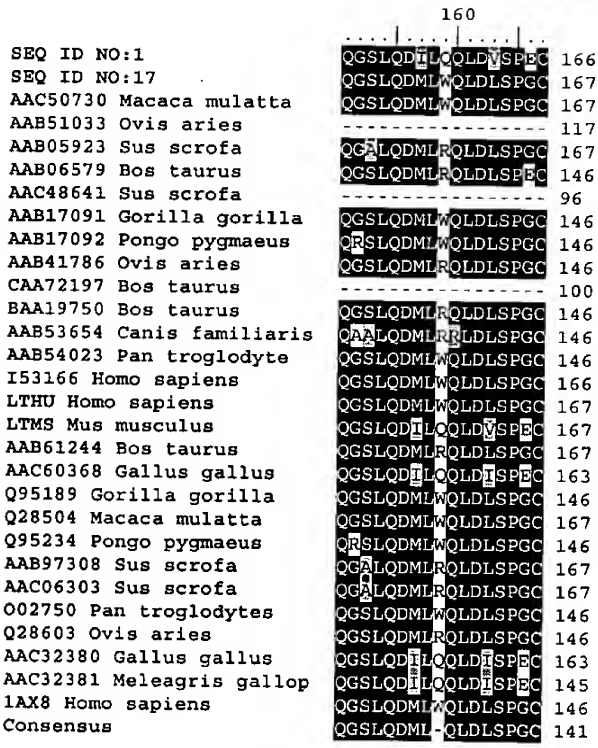
--33. (Twice Amended) [Any one of the peptides] The peptide of claim 1, wherein the peptide is cyclized. --

--61. (Amended) A purified peptide comprising the amino acid [residues] sequence of SEQ ID NO:18. --

Appendix C: ClustalW Alignment of Leptin Peptides Disclosed Before August 1999




110
120
130
140
150
SEQ ID NO:1
SEQ ID NO:17
AAC50730 Macaca mulatta
AAB51033 Ovis aries
AAB05923 Sus scrofa
AAB06579 Bos taurus
AAC48641 Sus scrofa
AAB17091 Gorilla gorilla
AAB17092 Pongo pygmaeus
AAB41786 Ovis aries
CAA72197 Bos taurus
BAA19750 Bos taurus
AAB53654 Canis familiaris
AAB54023 Pan troglodyte
I53166 Homo sapiens
LTHU Homo sapiens
LTMS Mus musculus
AAB61244 Bos taurus
AAC60368 Gallus gallus
Q95189 Gorilla gorilla
Q28504 Macaca mulatta
Q95234 Pongo pygmaeus
AAB97308 Sus scrofa
AAC06303 Sus scrofa
O02750 Pan troglodytes
Q28603 Ovis aries
AAC32380 Gallus gallus
AAC32381 Meleagris gallop
1AX8 Homo sapiens
Consensus
SEQ ID NO:18
SCHLPWA


160
SEQ ID NO:1
SEQ ID NO:17
AAC50730 Macaca mulatta
AAB51033 Ovis aries
AAB05923 Sus scrofa
AAB06579 Bos taurus
AAC48641 Sus scrofa
AAB17091 Gorilla gorilla
AAB17092 Pongo pygmaeus
AAB41786 Ovis aries
CAA72197 Bos taurus
BAA19750 Bos taurus
AAB53654 Canis familiaris
AAB54023 Pan troglodyte
I53166 Homo sapiens
LTHU Homo sapiens
LTMS Mus musculus
AAB61244 Bos taurus
AAC60368 Gallus gallus
Q95189 Gorilla gorilla
Q28504 Macaca mulatta
Q95234 Pongo pygmaeus
AAB97308 Sus scrofa
AAC06303 Sus scrofa
O02750 Pan troglodytes
Q28603 Ovis aries
AAC32380 Gallus gallus
AAC32381 Meleagris gallop
1AX8 Homo sapiens
Consensus

Quick Lookup

DICTIONARY

▸ LOOKUP

THESAURUS

▸ LOOKUP

▸ Add our Lookup Button to your browser

Pronunciation Key

\&\ as **a** and **u** in **a**b**u**t
\^\ as **e** in kitt**e**n
\&r\ as **ur/er** in f**ur**th**er**
\a\ as **a** in **a**sh
\A\ as **a** in **a**ce
\ä\ as **o** in m**o**p
\au\ as **ou** in **ou**t
\ch\ as **ch** in **ch**in
\e\ as **e** in b**e**t
\E\ as **ea** in **ea**sy
\g\ as **g** in **g**o
\i\ as **i** in h**i**t
\I\ as **i** in **i**ce
\j\ as **j** in **j**ob
\[ng]\ as **ng** in si**ng**
\O\ as **o** in **o**ld
\o\ as **aw** in **aw**l
\oi\ as **oy** in b**oy**
\th\ as **th** in **th**in
\[th]\ as **th** in **th**e
\ü\ as **oo** in l**oo**t
\u\ as **oo** in f**oo**t
\y\ as **y** in **y**et
\zh\ as **s**i in vi**s**ion

Dictionary Results

Merriam-Webster's COLLEGIATE® DICTIONARY

1 word found.

▸ THESAURUS

Main Entry: **ho.mo.logue**
Function: *noun*
Date: 1848
Variant(s): or ho.mo.log /'hO-m&-"log, 'h@-, -"l@g/
: something (as a chemical compound or a chromosome) homologous

▲ BACK TO TOP

Quick Lookup

DICTIONARY

▸ LOOKUP

THESAURUS

▸ LOOKUP

▸ Add our Lookup Button to your browser

Pronunciation Key

\&\ as **a** and **u** in **abut**
\ᵊ\ as **e** in kitten
\&r\ as **ur/er** in **further**
\a\ as **a** in **ash**
\A\ as **a** in **ace**
\ä\ as **o** in **mop**
\au\ as **ou** in **out**
\ch\ as **ch** in **chin**
\e\ as **e** in **bet**
\E\ as **ea** in **easy**
\g\ as **g** in **go**
\i\ as **i** in **hit**
\I\ as **i** in **ice**
\j\ as **j** in **job**
\[ng]\ as **ng** in **sing**
\O\ as **o** in **go**
\o\ as **aw** in **law**
\oi\ as **oy** in **boy**
\th\ as **th** in **thin**
\[th]\ as **th** in **the**
\ü\ as **oo** in **loot**
\u\ as **oo** in **foot**
\y\ as **y** in **yet**
\zh\ as **si** in **vision**

Dictionary Results

Merriam-Webster's COLLEGIATE® DICTIONARY

1 word found.

▸ THESAURUS

Main Entry: **ho.mol.o.gous**
Pronunciation: hO-'mä-l&-g&s, h&-
Function: *adjective*
Etymology: Greek homologos agreeing, from hom- + legein to say -- more at LEGEND
Date: 1660
1 a : having the same relative position, value, or structure: as (1) : exhibiting biological homology (2) : having the same or allelic genes with genetic loci usually arranged in the same order b : belonging to or consisting of a chemical series whose successive members have a regular difference in composition especially of one methylene group 2 : derived from or developed in response to organisms of the same species

▲ BACK TO TOP

Home | **The Dictionaries:** Language | Specialty | Multi-lingual | Translation | Other Indexes
'nyms & more | Grammars | Language Identifiers | Research | Library | Fun & Games | More
About yourDictionary.com | YD Lookup Button | Store | Register | Contact

Dictionary Results

Merriam-Webster's COLLEGIATE® DICTIONARY

3 words found.

To view an entry in the list, highlight it and click on GO TO.

▸ THESAURUS

▸ GOTO

an.a.log
an.a.logue[1,noun]
analog computer

Main Entry: **an.a.log**
Pronunciation: 'a-n[^&]l-"og, -"äg
Function: *adjective*
Date: 1948
1 : of, relating to, or being an analogue
2 a : of, relating to, or being a mechanism in which data is represented by continuously variable physical quantities b : of or relating to an analog computer c : being a timepiece having hour and minute hands

Main Entry: **[1]an.a.logue**
Function: *noun*
Etymology: French analogue, from analogue analogous, from Greek analogos
Date: 1826
Variant(s): or an.a.log /'a-n[^&]l-"og, -"@g/
1 : something that is analogous or similar to something else
2 : an organ similar in function to an organ of another animal or plant but different in structure and origin
3 usually analog : a chemical compound that is structurally similar to another but differs slightly in composition (as in the replacement of one atom by an atom of a different element or in the presence of a particular functional group)
4 : a food product made by combining a less expensive food (as soybeans or whitefish) with additives to give the appearance and taste of a more expensive food (as beef or crab)

Main Entry: **analog computer**
Function: *noun*
Date: 1948
: a computer that operates with numbers represented by directly measurable quantities (as voltages or rotations) -- compare DIGITAL COMPUTER, HYBRID COMPUTER

Pronunciation Key

\&\ as **a** and **u** in **abut**
\ᵊ\ as **e** in **kitten**
\&r\ as **ur/er** in **further**
\a\ as **a** in **ash**
\A\ as **a** in **ace**
\ä\ as **o** in **mop**
\au\ as **ou** in **out**
\ch\ as **ch** in **chin**
\e\ as **e** in **bet**
\E\ as **ea** in **easy**
\g\ as **g** in **go**
\i\ as **i** in **hit**
\I\ as **i** in **ice**
\j\ as **j** in **job**
\[ng]\ as **ng** in **sing**
\O\ as **o** in **go**
\o\ as **aw** in **law**
\oi\ as **oy** in **boy**
\th\ as **th** in **thin**
\[th]\ as **th** in **the**
\ü\ as **oo** in **loot**
\u\ as **oo** in **foot**
\y\ as **y** in **yet**
\zh\ as **si** in **vision**

▲ BACK TO TOP

Home | **The Dictionaries:** Language | Specialty | Multi-lingual | Translation | Other Indexes
'nyms & more | Grammars | Language Identifiers | Research | Library | Fun & Games | More
About yourDictionary.com | YD Lookup Button | Store | Register | Contact

Quick Lookup

DICTIONARY

▸ LOOKUP

THESAURUS

▸ LOOKUP

▸ Add our Lookup Button to your browser

Pronunciation Key

\&\ as **a** and **u** in **abut**
\ᵊ\ as **e** in kitt**e**n
\&r\ as **ur/er** in f**ur**th**er**
\a\ as **a** in **a**sh
\A\ as **a** in **a**ce
\ä\ as **o** in m**o**p
\au\ as **ou** in **ou**t
\ch\ as **ch** in **ch**in
\e\ as **e** in b**e**t
\E\ as **ea** in **ea**sy
\g\ as **g** in **g**o
\i\ as **i** in h**i**t
\I\ as **i** in **i**ce
\j\ as **j** in **j**ob
\[ng]\ as **ng** in si**ng**
\O\ as **o** in **o**ld
\o\ as **aw** in **aw**ful
\oi\ as **oy** in b**oy**
\th\ as **th** in **th**in
\[th]\ as **th** in **th**e
\ü\ as **oo** in l**oo**t
\u\ as **oo** in f**oo**t
\y\ as **y** in **y**et
\zh\ as **si** in vi**si**on

Dictionary Results

Merriam-Webster's COLLEGIATE® DICTIONARY

3 words found.

To view an entry in the list, highlight it and click on GO TO

▸ THESAURUS

▸ GOTO

de.riv.a.tive[1,noun]
derivative[2,adjective]
partial derivative

Main Entry: **[1]de.riv.a.tive**
Pronunciation: di-'ri-v&-tiv
Function: *noun*
Date: 15th century
1 : a word formed by derivation
2 : something derived
3 : the limit of the ratio of the change in a function to the corresponding change in its independent variable as the latter change approaches zero
4 a : a chemical substance related structurally to another substance and theoretically derivable from it b : a substance that can be made from another substance

Main Entry: **[2]derivative**
Function: *adjective*
Date: circa 1530
1 : formed by derivation
2 : made up of or marked by derived elements
3 : lacking originality : BANAL
- de.riv.a.tive.ly adverb
- de.riv.a.tive.ness noun

Main Entry: **partial derivative**
Function: *noun*
Date: 1889
: the derivative of a function of several variables with respect to one of them and with the remaining variables treated as constants

▲ BACK TO TOP

J. Mol. Biol. (1970) 48, 443–453

A General Method Applicable to the Search for Similarities in the Amino Acid Sequence of Two Proteins

Saul B. Needleman and Christian D. Wunsch

Department of Biochemistry, Northwestern University, and Nuclear Medicine Service, V. A. Research Hospital Chicago, Ill. 60611, U.S.A.

(*Received 21 July 1969*)

A computer adaptable method for finding similarities in the amino acid sequences of two proteins has been developed. From these findings it is possible to determine whether significant homology exists between the proteins. This information is used to trace their possible evolutionary development.

The maximum match is a number dependent upon the similarity of the sequences. One of its definitions is the largest number of amino acids of one protein that can be matched with those of a second protein allowing for all possible interruptions in either of the sequences. While the interruptions give rise to a very large number of comparisons, the method efficiently excludes from consideration those comparisons that cannot contribute to the maximum match.

Comparisons are made from the smallest unit of significance, a pair of amino acids, one from each protein. All possible pairs are represented by a two-dimensional array, and all possible comparisons are represented by pathways through the array. For this maximum match only certain of the possible pathways must be evaluated. A numerical value, one in this case, is assigned to every cell in the array representing like amino acids. The maximum match is the largest number that would result from summing the cell values of every pathway.

1. Introduction

The amino acid sequences of a number of proteins have been compared to determine whether the relationships existing between them could have occurred by chance. Generally, these sequences are from proteins having closely related functions and are so similar that simple visual comparisons can reveal sequence coincidence. Because the method of visual comparison is tedious and because the determination of the significance of a given result usually is left to intuitive rationalization, computer-based statistical approaches have been proposed (Fitch, 1966; Needleman & Blair, 1969).

Direct comparison of two sequences, based on the presence in both of corresponding amino acids in an identical array, is insufficient to establish the full genetic relationships between the two proteins. Allowance for gaps (Braunitzer, 1965) greatly multiplies the number of comparisons that can be made but introduces unnecessary and partial comparisons.

2. A General Method for Sequence Comparison

The smallest unit of comparison is a pair of amino acids, one from each protein. The maximum match can be defined as the largest number of amino acids of one protein that can be matched with those of another protein while allowing for all possible deletions.

The maximum match can be determined by representing in a two-dimensional array, all possible pair combinations that can be constructed from the amino acid sequences of the proteins, A and B, being compared. If the amino acids are numbered from the N-terminal end, Aj is the jth amino acid of protein A and Bi is the ith amino acid of protein B. The Aj represent the columns and the Bi the rows of the two-dimensional array, MAT. Then the cell, MATij, represents a pair combination that contains Aj and Bi.

Every possible comparison can now be represented by pathways through the array. An i or j can occur only once in a pathway because a particular amino acid cannot occupy more than one position at one time. Furthermore, if MATmn is part of a pathway including MATij, the only permissible relationships of their indices are $m > i, n \geqslant j$ or $m < i, n < j$. Any other relationships represent permutations of one or both amino acid sequences which cannot be allowed since this destroys the significance of a sequence. Then any pathway can be represented by MAT$ab \ldots$ MATyz, where $a \geqslant 1$, $b \geqslant 1$, the i and j of all subsequent cells of MAT are larger than the running indices of the previous cell and $y \leqslant K$, $z \leqslant M$, the total number of amino acids comprising the sequences of proteins A and B, respectively. A pathway is signified by a line connecting cells of the array. Complete diagonals of the array contain no gaps. When MATij and MATmn are part of a pathway, $i - m \neq j - n$ is a sufficient, but not necessary condition for a gap to occur. A necessary pathway through MAT is defined as one which begins at a cell in the first column or the first row. Both i and j must increase in value; either i or j must increase by only one but the other index may increase by one or more. This leads to the next cell in a MAT pathway. This procedure is repeated until either i or j, or both, equal their limiting values, K and M, respectively. Every partial or unnecessary pathway will be contained in at least one necessary pathway.

In the simplest method, MATij is assigned the value, one, if Aj is the same kind of amino acid as Bi; if they are different amino acids, MATij is assigned the value, zero. The sophistication of the comparison is increased if, instead of zero or one, each cell value is made a function of the composition of the proteins, the genetic code triplets representing the amino acids, the neighboring cells in the array, or any theory concerned with the significance of a pair of amino acids. A penalty factor, a number subtracted for every gap made, may be assessed as a barrier to allowing the gap. The penalty factor could be a function of the size and/or direction of the gap. No gap would be allowed in the operation unless the benefit from allowing that gap would exceed the barrier. The maximum-match pathway then, is that pathway for which the sum of the assigned cell values (less any penalty factors) is largest. MAT can be broken up into subsections operated upon independently. The method also can be expanded to allow simultaneous comparison of several proteins using the amino acid sequences of n proteins to generate an n-dimensional array whose cells represent all possible combinations of n amino acids, one from each protein.

The maximum-match pathway can be obtained by beginning at the terminals of the sequences ($i = y, j = z$) and proceeding toward the origins, first by adding to the value of each cell possessing indices $i = y - 1$ and/or $j = z - 1$, the maximum value from among all the cells which lie on a pathway to it. The process is repeated for indices $i = y - 2$ and/or $j = z - 2$. This increment in the indices is continued until all cells in the matrix have been operated upon. Each cell in this outer row or column will contain the maximum number of matches that can be obtained by originating

any pathway at that cell and the largest number in that row or column is equal to the maximum match; the maximum-match pathway in any row or column must begin at this number. The operation of successive summations of cell values is illustrated in Figures 1 and 2.

	A	B	C	N	J	R	Q	C	L	C	R	P	M
A	1												
J					1								
C			1					1		1			
J					1								
N				1									
R						1	4	3	3	2	2	0	0
C	3	3	4	3	3	3	3	4	3	3	1	0	0
K	3	3	3	3	3	3	3	3	3	2	1	0	0
C	2	2	3	2	2	2	2	3	2	3	1	0	0
R	2	1	1	1	1	2	1	1	1	1	2	0	0
B	1	2	1	1	1	1	1	1	1	1	1	0	0
P	0	0	0	0	0	0	0	0	0	0	0	1	0

FIG. 1. The maximum-match operation for necessary pathways.
The number contained in each cell of the array is the largest number of identical pairs that can be found if that cell is the origin for a pathway which proceeds with increases in running indices. Identical pairs of amino acids were given the value of one. Blank cells which represent non-identical pairs have the value, zero. The operation of successive summations was begun at the last row of the array and proceeded row-by-row towards the first row. The operation has been partially completed in the R row. The enclosed cell in this row is the site of the cell operation which consists of a search along the subrow and subcolumn indicated by borders for the largest value, 4 in subrow C. This value is added to the cell from which the search began.

	A	B	C	N	J	R	O	C	L	C	R	P	M
A	8	7	6	6	5	4	4	3	3	2	1	0	0
J	7	7	6	6	6	4	4	3	3	2	1	0	0
C	6	6	7	6	5	4	4	4	3	3	1	0	0
J	6	6	6	5	6	4	4	3	3	2	1	0	0
N	5	5	5	6	5	4	4	3	3	2	1	0	0
R	4	4	4	4	4	5	4	3	3	2	2	0	0
C	3	3	4	3	3	3	3	4	3	3	1	0	0
K	3	3	3	3	3	3	3	3	3	2	1	0	0
C	2	2	3	2	2	2	2	3	2	3	1	0	0
R	2	1	1	1	1	2	1	1	1	1	2	0	0
B	1	2	1	1	1	1	1	1	1	1	1	0	0
P	0	0	0	0	0	0	0	0	0	0	0	1	0

FIG. 2. Contributors to the maximum match in the completed array.
The alternative pathways that could form the maximum match are illustrated. The maximum match terminates at the largest number in the first row or first column, 8 in this case.

It is apparent that the above array operation can begin at any of a number of points along the borders of the array, which is equivalent to a comparison of N-terminal residues or C-terminal residues only. As long as the appropriate rules for pathways are followed, the maximum match will be the same. The cells of the array which contributed to the maximum match, may be determined by recording the origin of the number that was added to each cell when the array was operated upon.

3. Evaluating the Significance of the Maximum Match

A given maximum match may represent the maximum number of amino acids matched, or it may just be a number that is a complex function of the relationship between sequences. It will, however, always be a function of both the amino acid compositions of the proteins and the relationship between their sequences. One may ask whether a particular result found differs significantly from a fortuitous match between two random sequences. Ideally, one would prefer to know the exact probability of obtaining the result found from a pair of random sequences and what fraction of the total possibilities are less probable, but that is prohibitively difficult, especially if a complex function were used for assigning a value to the cells.

As an alternative to determining the exact probabilities, it is possible to estimate the probabilities experimentally. To accomplish the estimate one can construct two sets of random sequences, a set from the amino acid composition of each of the proteins compared. Pairs of random sequences can then be formed by randomly drawing one member from each set. Determining the maximum match for each pair selected will yield a set of random values. If the value found for the real proteins is significantly different from the values found for the random sequences, the difference is a function of the sequences alone and not of the compositions. Alternatively, one can construct random sequences from only one of the proteins and compare them with the other to determine a set of random values. The two procedures measure different probabilities. The first procedure determines whether a significant relationship exists *between* the real sequences. The second procedure determines whether the relationship *of* the protein used to form the random sequences *to* the other proteins is significant. It bears reiterating that the integral amino acid composition of each random sequence must be equal to that of the protein it represents.

The amino acid sequence of each protein compared belongs to a set of sequences which are permutations. Sequences drawn randomly from one or both of these sets are used to establish a distribution of random maximum-match values which would include all possible values if enough comparisons were made. The null hypothesis, that any sequence relationship manifested by the two proteins is a random one, is tested. If the distribution of random values indicates a small probability that a maximum match equal to, or greater than, that found for the two proteins could be drawn from the random set, the hypothesis is rejected.

4. Cell Values and Weighting Factors

To provide a theoretical framework for experiments, amino acid pairs may be classified into two broad types, identical and non-identical pairs. From 20 different amino acids one can construct 180 possible non-identical pairs. Of these, 75 pairs of amino acids have codons (Marshall, Caskey & Nirenberg, 1967) whose bases differ in only one position (Eck & Dayhoff, 1966). Each change is presumably the result of

single-point mutation. The majority of non-identical pairs have a maximum of only one or zero corresponding bases. Due to the degeneracy of the genetic code, pair differences representing amino acids with no possible corresponding bases are uncommon even in randomly selected pairs. If cells are weighted in accordance with the maximum number of corresponding bases in codons of the represented amino acids, the maximum match will be a function of identical and non-identical pairs. For comparisons in general, the cell weights can be chosen on any basis.

If every possible sequence gap is allowed in forming the maximum match, the significance of the maximum match is enhanced by decreasing the weight of those pathways containing a large number of gaps. A simple way to accomplish this is to assign a penalty factor, a number which is subtracted from the maximum match for each gap used to form it. The penalty is assigned before the maximum match is formed. Thus the pathways will be weighted according to the number of gaps they contain, but the nature of the contributors to the maximum match will be affected as well. In proceeding from one cell to the next in a maximum-match pathway, it is necessary that the difference between each cell value and the penalty, be greater than the value for a cell in a pathway that contains no gap. If the value of the penalty were zero, all possible gaps could be allowed. If the value were equal to the theoretical value for the maximum match between two proteins, it would be impossible to allow a gap and the maximum match would be the largest of the values found by simply summing along the diagonals of the array; this is the simple frame-shift method.

5. Application of the Method

To illustrate the role of weighting factors in evaluating a maximum match, two proteins expected to show homology, whale myoglobin (Edmundson, 1965) and human β-hemoglobin (Konigsberg, Goldstein & Hill, 1963), and two proteins not expected to exhibit homology, bovine pancreatic ribonuclease (Smyth, Stein & Moore, 1963) and hen's egg lysozyme (Canfield, 1963) were chosen for comparisons.

The FORTRAN programs used in this study were written for the CDC3400 computer. The operations employed in forming the maximum match are those for the special case when none of the cells of the array have a value less than zero. Four types of amino acid pairs were distinguished and variable sets consisting of values to be assigned to each type of pair and a value for the penalty were established. The pair types are as follows:

Type 3. Identical pairs: those having a maximum of three corresponding bases in their codons.

Type 2. Pairs having a maximum of two corresponding bases in their codons.

Type 1. Pairs having a maximum of one corresponding base in their codons.

Type 0. Pairs having no possible corresponding base in their codons.

The value for type 3 pairs was 1·0 and the value for type 0 pairs was zero for all variable sets.

At program execution time, the amino acids (coded by two-digit numbers) of the sequences to be compared were read into the computer, and were followed by a twenty-by-twenty symmetrical array, the maximum correspondence array, analogous to one used by Fitch (1966), that contained all possible pairs of amino acids and identified each pair as to type. The RNA codons for amino acids used to construct the maximum-correspondence array were taken from a single Table (Marshall *et al.*,

1967). The UGA, UAA and UAG codons were not used, but UUG was used as a codon for leucine. The subsequent data cards indicated the numerical values for a variable set.

The two-dimensional comparison array was generated row-by-row. The amino acid code numbers for Ai and Bj referenced the correspondence array to determine the type of amino acid pair constituted by Ai and Bj. The type number referenced a third array, the variable set, containing the type values, and the appropriate value from that set was assigned to the appropriate cell of the comparison array. The maximum match was then determined by the procedure of successive summations.

Following the determination of the maximum match for the real proteins, the amino acid sequence of only one member of the protein pair was randomized and the match was repeated. The sequences of β-hemoglobin and ribonuclease were the ones randomized. The randomization procedure was a sequence shuffling routine based on computer-generated random numbers. A cycle of sequence randomization–maximum-match determination was repeated ten times in all of the experiments in this report, giving the random values used for comparison with the real maximum-match. The average and standard deviation for the random values of each variable set was estimated.

6. Results and Discussion

The use of a small random sample size (ten) was necessary to hold the computer time to a reasonable level. The maximum probable error in a standard deviation estimate for a sample this small is quite large and the results should be judged with this fact in mind. For each set of variables, it was assumed that the random values would be distributed in the fashion of the normal-error curve; therefore, the values of the first six random sets in the β-hemoglobin–myoglobin comparison were converted to standard measure, five was added to the result, and these values were plotted as *one* group against their calculated probit. The results of the plot are shown in Figure 3. The fit is good indicating the probable adequacy of the measured standard deviations for these variable sets in estimating distribution functions for random values through two standard deviations. The above fit indicates no bias in the randomization procedure. In other words, randomization of the sequence was complete before the maximum match was determined for any sequence in a random set.

The results obtained in the comparison of β-hemoglobin with myoglobin are summarized in Table 1 and the results for the ribonuclease–lysozyme comparison are in Table 2. These Tables indicate the values assigned to the pair types, the penalty factor used in forming each of the maximum matches, and the statistical results obtained. The number of gaps roughly characterizes the nature of the pathway that formed the maximum match. A large number is indicative of a devious pathway through the array. One gap means that all of the pathway may be found on only two partial diagonals of the array.

The most important information is obtained from the standardized value of the maximum match for the real proteins, the difference from the mean in standard deviation units. For this sample size all deviations greater than 3·0 were assumed to include less than 1% of the true random population and to indicate a significant difference. As might be expected, all matches of myoglobin and β-hemoglobin show a significant deviation. Among the sets of variables, set 1, which results in search for identical amino acid pairs while allowing for all deletions, indicates that

The solid li
of samples fr
calculations o
population.

Variable set	M
1	
2	
3	0·
4	0·
5	0·
6	0·
7	0

e is the esti
mum match o
n each varia
An avera

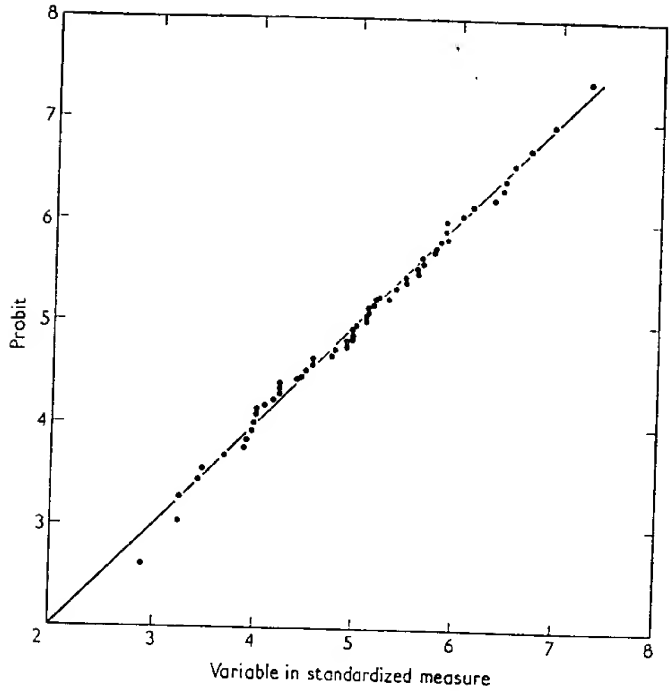


FIG. 3. Probit plot for six grouped random samples.
The solid line indicates the plot that would result from a probit analysis on an infinite number of samples from a normally-distributed population. The points represent the results of probit calculations on 60 random maximum match values that were assumed to have come from one population.

TABLE 1

β-Hemoglobin–myoglobin maximum matches

Variable set	Match values for pair types 2	1	Penalty	Maximum-match value sum Real	Random†	s	Real X	Minimum deletions Real	Random†
1	0	0	0	63·00	55·60	1·80	4·11	35	36·2
2	0	0	1·00	38·00	27·80	2·09	4·88	4	5·5
3	0·67	0·33	0	97·00	91·47	1·55	3·57	18	24·3
4	0·67	0·33	1·03	89·63	80·25	1·11	8·46	1	3·6
5	0·25	0·05	0	71·55	64·78	1·59	4·27	46	45·0
6	0·25	0·05	1·05	51·95	40·54	1·46	7·80	3	7·5
7	0·25	0·05	25	47·30	33·80	1·52	8·87	0	0

s is the estimated standard deviation; X, the standardized value, (real − random)/s, of the maximum match of the real proteins. The values for type 3 and type 0 pairs were 1·0 and 0, respectively, in each variable set.

† An average value from 10 samples.

TABLE 2

Ribonuclease–lysozyme maximum matches

Variable set	Match values for pair types		Penalty	Maximum-match value sum		s	Real X	Minimum deletions	
	2	1		Real	Random†			Real	Random†
1	0	0	0	48·00	44·20	2·56	1·48	34	29·2
2	0	0	1·00	23·00	22·00	1·73	0·58	5	5·2
3	0·67	0·33	0	78·33	76·17	0·82	2·64	21	18·3
4	0·67	0·33	1·03	67·93	67·37	1·27	0·43	2	2·2
5	0·25	0·05	0	56·00	52·26	2·12	1·77	35	35·5
6	0·25	0·05	1·05	33·70	33·02	1·66	0·41	8	6·8
7	0·25	0·05	25	28·15	27·67	1·75	0·22	0	0

s is the estimated standard deviation; X, the standardized value, (real − random)/s, of the maximum match of the real proteins. The values for type 3 and type 0 pairs were 1·0 and 0, respectively in each variable set.

† An average value from 10 samples.

amino acids in β-hemoglobin and myoglobin can be matched. To attain this match, however, it is necessary to permit at least 35 gaps. In contrast, when two gaps are allowed according to Braunitzer (1965), it is possible to match only 37 of the amino acids. Curiously, when this variable set was used for comparing *human* myoglobin (Hill, personal communication) with human β-hemoglobin, the maximum match obtained was not significant. Differences between real and random values were highly significant, however, when other variable sets were used.

Variable set 2 attaches a penalty equal to the value of one identical amino acid pair to the search for identical amino acid pairs. This penalty will exclude from consideration any possible pathway that leaves and returns to a principal diagonal, thereby needing two gaps, in order to add only one or two amino acids to the maximum match. This set results in a total of $30 + 4 = 42$ amino acids matched (the maximum-match value plus the number of gaps is reduced to four) and the significance of the result relative to set 1 appears to be increased. Braunitzer's comparison would have a value of $37 - 2 = 35$ using this variable set, hence it was not selected by the method.

Variable sets 3 and 4 have an interesting property. Their maximum-match values can be related to the minimum number of mutations needed to convert the selected parts of one amino acid sequence into the selected parts of the other. The minimum number of mutations concept in protein comparisons was first advanced by Fitch (1966). If the type values for these sets are multiplied by three, they become equal to their pair type and directly represent the maximum number of corresponding bases in the codons for a given amino acid pair. Thus the maximum match and penalty factors may be multiplied by three, making it possible to calculate the maximum number of bases matched in the combination of amino acid pairs selected by the maximum-match operation.

β-Hemoglobin, the smaller of the two proteins, contains 146 amino acids; consequently the highest possible maximum match (disregarding integral amino-acid composition data) with myoglobin is $146 \times 3 = 438$. Insufficient data are avail-

to analyze the result from set 3 on the basis of mutations. If it is assumed that the gap in set 4 does not exclude any part of β-hemoglobin from the comparison, this set has a maximum of 3(89·63 + 1·03) = 272 bases matched, indicating a minimum of 438 − 272 = 166 point mutations in this combination. Using this variable set and placing gaps according to Braunitzer, a score of 88·6 was obtained, thus his match was not selected. Again it may be observed that the penalty greatly enhanced the significance of the maximum match.

Variable sets 5 and 6 have no intrinsic meaning and were chosen because the weight attached to type 2 and type 1 pairs is intermediate in value with respect to sets 1 and 2 and sets 3 and 4. The maximum match for set 6 is seen to have a highly significant value.

The data of set 7 are results that would be obtained from using the frame-shift method to select a maximum match; the penalty was large enough to prevent any gaps in the comparisons. The slight differences in significance found among the maximum-match values of β-hemoglobin and myoglobin resulting from use of sets 4, 6 and 7 are probably meaningless due to small sample size and errors introduced by the assumptions about the distribution functions of random values. Finding a value in set 7 that is approximately equal to those from sets 4 and 6 in significance is not surprising. A larger penalty factor would have increased the difference from the mean in sets 4 and 6 because almost every random value in each set was the result of more gaps than were required to form the real maximum match. Further, the gaps that are allowed are at the N-terminal ends so that about 85% of the comparison can be made without gaps. If an actual gap were present near the middle of one of the sequences, it would have caused a sharp reduction in the significance of the frame-shift type of match.

Set 3 is the only variable set in Table 2 that shows a possible difference. Assuming the value is accurate, other than chance, there is no simple explanation for the difference. A small but meaningful difference in any comparison could represent evolutionary divergence or convergence. It is generally accepted that the primary structure of proteins is the chief determinant of the tertiary structure. Because certain features of tertiary structure are common to proteins, it is reasonable to suppose that proteins will exhibit similarities in their sequences, and that these similarities will be sufficient to cause a significant difference between most protein pairs and their corresponding randomized sequences, being an example of submolecular evolutionary convergence. Further, the interactions of the protein backbone, side chains, and the solvent that determine tertiary structure are, in large measure, forces arising from the polarity and steric nature of the protein side-chains. There are conspicuous correlations in the polarity and steric nature of type 2 pairs. Heavy weighting of these pairs would be expected to enhance the significance of real maximum-match values if common structural features are present in proteins that are compared. The presence of sequence similarities does not always imply common ancestry in proteins. More experimentation will be required before a choice among the possibilities suggested for the result from set 3 can be made. If several short sequences of amino acids are common to all proteins, it seems remarkable that the relationship of ribonuclease to lysozyme in six of the seven variable sets appears to be truly a random one. It should be noted, however, that the standard value of the real maximum-match is positive in each variable set in this comparison.

This method was designed for the purpose of detecting homology and defining its nature when it can be shown to exist. Its usefulness for the above purposes depends in

part upon assumptions related to the genetic events that could have occurred in the evolution of proteins. Starting with the assumption that homologous proteins are the result of gene duplication and subsequent mutations, it is possible to construct several hypothetical amino-acid sequences that would be expected to show homology. If one assumes that following the duplication, point mutations occur at a constant or variable rate, but randomly, along the genes of the two proteins, after a relatively short period of time the protein pairs will have nearly identical sequences. Detection of the high degree of homology present can be accomplished by several means. The use of values for non-identical pairs will do little to improve the significance of the results. If no, or very few, deletions (insertions) have occurred, one could expect to enhance the significance of the match by assigning a relatively high penalty for gaps. Later on in time the hypothetical proteins may have a sizable fraction of their codons changed by point mutations, the result being that an attempt to increase the significance of the maximum match will probably require attaching substantial weight to those pairs representing amino acids still having two of the three original bases in their codons. Further, if a few more gaps have occurred, the penalty should be reduced to a small enough value to allow areas of homology to be linked to one another. At a still later date in time more emphasis must be placed on non-identical pairs, and perhaps a very small or even negative penalty factor must be assessed. Eventually, it will be impossible to detect the remaining homology in the hypothetical example by using the approach detailed here.

From consideration of this simple model of protein evolution one may deduce that the variables which maximize the significance of the difference between real and random proteins gives an indication of the nature of the homology. In the comparison of human β-hemoglobin to whale myoglobin, the assignment of some weight to type 2 pairs considerably enhances the significance of the result, indicating substantial evolutionary divergence. Further, few deletions (additions) have apparently occurred.

It is known that the evolutionary divergence manifested by cytochrome (Margoliash, Needleman & Stewart, 1963) and other heme proteins (Zuckerkandl & Pauling, 1965) did not follow the sample model outlined above. Their divergence is the result of *non-random* mutations along the genes. The degree and type of homology can be expected to differ between protein pairs. As a consequence of the difference there is no *a priori* best set of cell and operation values for maximizing the significance of a maximum-match value of homologous proteins, and as a corollary to this fact, there is no best set of values for the purpose of detecting only slight homology. This is an important consideration, because whether the sequence relationship between proteins is significant depends solely upon the cell and operation values chosen. If it is found that the divergence of proteins follows one or two simple models, it may be possible to derive a set of values that will be most useful in detecting and defining homology.

The most common method for determining the degree of homology between protein pairs has been to count the number of non-identical pairs (amino acid replacements) in the homologous comparison and to use this number as a measure of evolutionary distance between the amino acid sequences. A second, more recent concept has been to count the minimum number of mutations represented by the non-identical pairs. This number is probably a more adequate measure of evolutionary distance because it utilizes more of the available information and theory to give some measure of the number of genetic events that have occurred in the evolution of the proteins. The approach outlined in this paper supplies either of these numbers.

This work was supported in part by grants to one of us (S.B.N.) from the U.S. Public Health Service (1 501 FR 05370 02) and from Merck Sharp & Dohme.

REFERENCES

Braunitzer, G. (1965). In *Evolving Genes and Proteins*, ed. by V. Bryson & H. J. Vogel, p. 183. New York: Academic Press.
Canfield, R. (1963). *J. Biol. Chem.* **238**, 2698.
Eck, R. V. & Dayhoff, M. O. (1966). *Atlas of Protein Sequence and Structure*. Silver Spring, Maryland: National Biomedical Research Foundation.
Edmundson, A. B. (1965). *Nature*, **205**, 883.
Fitch, W. (1966). *J. Mol. Biol.* **16**, 9.
Konigsberg, W., Goldstein, J. & Hill, R. J. (1963). *J. Biol. Chem.* **238**, 2028.
Margoliash, E., Needleman, S. B. & Stewart, J. W. (1963). *Acta Chem. Scand.* **17**, S 250.
Marshall, R. E., Caskey, C. T. & Nirenberg, M. (1967). *Science*, **155**, 820.
Needleman, S. B. & Blair, T. H. (1969). *Proc. Nat. Acad. Sci., Wash.* **63**, 1127.
Smyth, D. G., Stein, W. G. & Moore, S. (1963). *J. Biol. Chem.* **238**, 227.
Zuckerkandl, E. & Pauling, L. (1965). In *Evolving Genes and Proteins*, ed. by V. Bryson & H. J. Vogel, p. 97. New York: Academic Press.

Quick Lookup

DICTIONARY

mammalian

▸ LOOKUP

THESAURUS

▸ LOOKUP

▸ Add our Lookup Button to your browser

Pronunciation Key

\&\ as **a** and **u** in **abut**
\ᵊ\ as **e** in **kitten**
\&r\ as **ur/er** in **further**
\a\ as **a** in **ash**
\A\ as **a** in **ace**
\ä\ as **o** in **mop**
\au\ as **ou** in **out**
\ch\ as **ch** in **chin**
\e\ as **e** in **bet**
\E\ as **ea** in **easy**
\g\ as **g** in **go**
\i\ as **i** in **hit**
\I\ as **i** in **ice**
\j\ as **j** in **job**
\[ng]\ as **ng** in **sing**
\O\ as **o** in **go**
\o\ as **aw** in **law**
\oi\ as **oy** in **boy**
\th\ as **th** in **thin**
\[th]\ as **th** in **the**
\ü\ as **oo** in **loot**
\u\ as **oo** in **foot**
\y\ as **y** in **yet**
\zh\ as **si** in **vision**

Dictionary Results

Merriam-Webster's COLLEGIATE® DICTIONARY

1 word found.

▸ THESAURUS

Main Entry: **mam.mal**
Pronunciation: 'ma-m&l
Function: *noun*
Etymology: New Latin Mammalia, from Late Latin, neuter plural of mammalis of the breast, from Latin mamma breast
Date: 1826
: any of a class (Mammalia) of warm-blooded higher vertebrates (as placentals, marsupials, or monotremes) that nourish their young with milk secreted by mammary glands, have the skin usually more or less covered with hair, and include humans
- mam.ma.li.an /m&-'mA-lE-&n, ma-/ adjective or noun

▲ BACK TO TOP

Quick Lookup

DICTIONARY

human

▸ LOOKUP

THESAURUS

▸ LOOKUP

▸ Add our Lookup Button to your browser

Pronunciation Key

\&\ as **a** and **u** in abut
\ᵊ\ as **e** in kitten
\&r\ as **ur/er** in further
\a\ as **a** in ash
\A\ as **a** in ace
\ä\ as **o** in mop
\au\ as **ou** in out
\ch\ as **ch** in chin
\e\ as **e** in bet
\E\ as **ea** in easy
\g\ as **g** in go
\i\ as **i** in hit
\I\ as **i** in ice
\j\ as **j** in job
\[ng]\ as **ng** in sing
\O\ as **o** in go
\o\ as **aw** in law
\oi\ as **oy** in boy
\th\ as **th** in thin
\[th]\ as **th** in the
\ü\ as **oo** in loot
\u\ as **oo** in foot
\y\ as **y** in yet
\zh\ as **si** in vision

Dictionary Results

Merriam-Webster's COLLEGIATE® DICTIONARY

2 words found.

To view an entry in the list, highlight it and click on GO TO.

▸ THESAURUS

▸ GOTO

mu.rine
murine typhus

Main Entry: **mu.rine**
Pronunciation: 'myur-"In
Function: *adjective*
Etymology: ultimately from Latin mur-, mus
Date: 1607
: of or relating to a murid genus (Mus) or its subfamily (Murinae) which includes the common household rats and mice; also : of, relating to, or involving these rodents and especially the house mouse
- murine noun

Main Entry: **murine typhus**
Function: *noun*
Date: 1933
: a mild febrile disease that is marked by headache and rash, is caused by a rickettsia (Rickettsia mooseri), is widespread in nature in rodents, and is transmitted to humans by a flea

▲ BACK TO TOP

Quick Lookup

DICTIONARY

▸ LOOKUP

THESAURUS

▸ LOOKUP

▸ Add our Lookup Button to your browser

Pronunciation Key

\&\ as **a** and **u** in abut
\ᵊ\ as **e** in kitten
\&r\ as **ur/er** in **further**
\a\ as **a** in ash
\A\ as **a** in ace
\ä\ as **o** in mop
\au\ as **ou** in **out**
\ch\ as **ch** in **ch**in
\e\ as **e** in bet
\E\ as **ea** in **easy**
\g\ as **g** in **go**
\i\ as **i** in hit
\I\ as **i** in ice
\j\ as **j** in **j**ob
\[ng]\ as **ng** in sing
\O\ as **o** in **go**
\o\ as **aw** in **law**
\oi\ as **oy** in boy
\th\ as **th** in **th**in
\[th]\ as **th** in **th**e
\ü\ as **oo** in l**oo**t
\u\ as **oo** in f**oo**t
\y\ as **y** in **y**et
\zh\ as **si** in vi**si**on

Dictionary Results

Merriam-Webster's COLLEGIATE® DICTIONARY

10 words found.

To view an entry in the list, highlight it and click on GO TO.

▸ THESAURUS

▸ GOTO

hu.man[1,adjective]
human[2,noun]
human being
human ecology
human engineering

Main Entry: **[1]hu.man**
Pronunciation: 'hyü-m&n, 'yü-
Function: *adjective*
Etymology: Middle English humain, from Middle French, from Latin humanus; akin to Latin homo human being -- more at HOMAGE
Date: 14th century
1 : of, relating to, or characteristic of humans
2 : consisting of humans
3 a : having human form or attributes b : susceptible to or representative of the sympathies and frailties of human nature
- hu.man.ness /-m&n-n&s/ noun

Main Entry: **[2]human**
Function: *noun*
Date: circa 1533
: a bipedal primate mammal (Homo sapiens) : MAN; broadly : any living or extinct member of the family (Hominidae) to which the primate belongs
- hu.man.like /-m&n-"lIk/ adjective

Main Entry: **human being**
Function: *noun*
Date: 1795
: HUMAN

Main Entry: **human ecology**
Function: *noun*
Date: 1907
1 : a branch of sociology dealing especially with the spatial and temporal

interrelationships between humans and their economic, social, and political organization
2 : the ecology of human communities and populations especially as concerned with preservation of environmental quality (as of air or water) through proper application of conservation and civil engineering practices

Main Entry: **human engineering**
Function: *noun*
Date: 1920
1 : management of humans and their affairs especially in industry
2 : ERGONOMICS

Main Entry: **human immunodeficiency virus**
Function: *noun*
Date: 1986
: HIV

Main Entry: **human nature**
Function: *noun*
Date: 1668
: the nature of humans; especially : the fundamental dispositions and traits of humans

Main Entry: **human relations**
Function: *noun plural but usually singular in construction*
Date: 1946
1 : a study of human problems arising from organizational and interpersonal relations (as in industry)
2 : a course, study, or program designed to develop better interpersonal and intergroup adjustments

Main Entry: **human resources**
Function: *noun plural*
Date: 1975
: PERSONNEL 1a, 2

Main Entry: **human rights**
Function: *noun plural*
Date: 1791
: rights (as freedom from unlawful imprisonment, torture, and execution) regarded as belonging fundamentally to all persons

▲ BACK TO TOP

Dictionary Results

Merriam-Webster's COLLEGIATE® DICTIONARY

5 words found.

To view an entry in the list, highlight it and click on GO TO.

▸ THESAURUS

▸ GOTO

- syn.thet.ic[1,adjective]
- synthetic[2,noun]
- synthetic division
- synthetic geometry
- synthetic resin

Main Entry: **[1]syn.thet.ic**
Pronunciation: sin-'the-tik
Function: *adjective*
Etymology: Greek synthetikos of composition, component, from syntithenai to put together
Date: 1697
1 : relating to or involving synthesis : not analytic
2 : attributing to a subject something determined by observation rather than analysis of the nature of the subject and not resulting in self-contradiction if negated -- compare ANALYTIC
3 : characterized by frequent and systematic use of inflected forms to express grammatical relationships
4 a (1) : of, relating to, or produced by chemical or biochemical synthesis; especially : produced artificially (2) : of or relating to a synfuel b : devised, arranged, or fabricated for special situations to imitate or replace usual realities c : FACTITIOUS, BOGUS
- syn.thet.i.cal.ly /-ti-k(&-)lE/ adverb

Main Entry: **[2]synthetic**
Function: *noun*
Date: 1916
: something resulting from synthesis rather than occurring naturally; especially : a product (as a drug or plastic) of chemical synthesis

Main Entry: **synthetic division**
Function: *noun*
Date: 1904
: a simplified method for dividing a polynomial by another polynomial of the first degree by writing down only the coefficients of the several

powers of the variable and changing the sign of the constant term in the divisor so as to replace the usual subtractions by additions

Main Entry: **synthetic geometry**
Function: *noun*
Date: 1889
: elementary euclidean geometry or projective geometry as distinguished from analytic geometry

Main Entry: **synthetic resin**
Function: *noun*
Date: 1907
: RESIN 2

▲ BACK TO TOP

Home | **The Dictionaries:** Language | Specialty | Multi-lingual | Translation | Other Indexes
'nyms & more | Grammars | Language Identifiers | Research | Library | Fun & Games | More
About yourDictionary.com | YD Lookup Button | Store | Register | Contact